A Graphical Exploration of Non-uniform Errors

Laura B. James†, Andrew W. Moore*, Madeleine Glick‡, Adrian Wonfor†
†University of Cambridge, Department of Engineering, Centre for Photonic Systems
{lbj20, aw300}@eng.cam.ac.uk
*University of Cambridge, Computer Laboratory, andrew.moore@cl.cam.ac.uk
‡Intel Research, Cambridge, madeleine.glick@intel.com

***Abstract*— We consider the propagation of errors through the network stack when a system is carrying real traffic, and the methods by which these errors are measured. Using Gigabit Ethernet on fibre in a state of reduced receiver power as an example, we find that bit error rate and packet error rate have only a weakly deterministic relationship. To illustrate the non-intuitive nature of the effects observed, we examine how physical layer errors interact with the line coding scheme used to create data-dependent error patterns.**

I. Introduction

Researchers and engineers often test optical components and sub-systems using pseudo-random bit sequences, and base their performance evaluations of these overall systems on the bit error rate (BER) thus obtained. However, networks carry real traffic very different in character to these artificial test sequences, thus making the frame loss, data integrity, and other high level network metrics more important to the network's operators and users than BER.

We describe here some of the effects of errors on data at various levels in the network stack. In particular, we consider the ways in which physical layer errors manifest themselves in user data when a popular line coding scheme is used.

Optical networking motivation

Current work in all areas of networking has led to increasingly complex architectures: our interest is focused upon the field of optical networking. Our exploration of the robustness of network systems is motivated by the increased demands of these new optical systems. To take advantage of capacity developments offered by optical systems at the short timescales relevant to local area networks, packet switching and burst switching techniques have seen significant investigation. An example is our project to investigate Optical Packet Switching, involving the construction of a switched optical data path based upon semiconductor optical amplifiers for use in system- and local-area network situations [1]. This work attempts to minimise latency, and avoids both optical buffering and all-optical signal processing to ensure lowest cost. Our architecture uses high speed optical switch fabrics for high speed routing, and combines this with wavelength striping and a separate control channel. Despite designing the system for short link lengths, the data path between the sending and receiving end-systems is clearly non-trivial, with a significant number of devices such as amplifiers and wavelength multiplex units in the path.

Deployments with longer runs of fibre may use large numbers of splitters for measurement and monitoring, as well as active optical devices; the overall system loss in these new, complex systems is therefore greater than in today's point-to-point links. Other examples include Ethernet in the last mile, and passive optical networks [2].

In contrast to the need for networks to operate at reduced power exists the desire for faster networks. If all other variables are held constant, an increase in bit rate will require a proportional increase in transmitter power to maintain the same bit error rate.

II. Bit Error Rate versus Packet Error Experiments

We chose Gigabit Ethernet on optical fibre [3] as the basis for our work. This uses the 8B/10B codec, originally described by Widmer & Franaszek [4], which converts 8 bits of data for transmission (ideal for any octet-orientated system) into a 10 bit line code. This codec is widely used and is also found in Fibre Channel, the 800Mbps extensions to the IEEE 1394 / Firewire standard, and is the basis of coding for the electrical signals of PCI Express.

We investigate Gigabit Ethernet under conditions where the received power is sufficiently low as to induce errors in the Ethernet frames. We assume that while the Functional Redundancy Check (FRC) mechanism within Ethernet is sufficiently strong to catch the errors, the dropped frames and resulting packet loss will result in a significantly higher probability of packet errors than the norm for certain hosts, applications and perhaps users.

In our main test environment an optical attenuator is placed in one direction of a Gigabit Ethernet link, between a traffic generator and a traffic sink and tester. A packet capture and measurement system is implemented in the traffic sink which allows application processes to receive errored frames that would normally be discarded. Pre-constructed test data in tcpdump format is transmitted from the generator, and purpose-built code compares transmitted frames to their received versions, and if they differ, stores both original and errored frames for octet-by-octet analysis.

Some results presented here are conducted with real network traffic referred to as the *day-trace*. This network traffic was captured from the interconnect between a large research institution and the Internet over the course of two working days. Other traffic tested included *pseudo-random data*, consisting

of a sequence of frames of the same number and size as the *day-trace* data, although each is filled with a stream of octets whose values were drawn from a pseudo-random number generator.

Motivating Results: We noted how errors are position independent but dependent upon the encoded data [5], and that certain payload octets display significantly higher error-rates. It is important to note that the errors occur uniformly across the whole packet and that there are no correlations evident between the positions of errors within the frame. We interpret this result as confirming that errors are highly localised within a frame and from this we are able to assume that the error-inducing events occur over small (bit-time) time scales. Frames containing different data lead to substantially different bit error rates, as well as different packet error performance. We further found the relationship between bit error rate and packet error rate was only weakly deterministic, and the relationship for the same input data was inconsistent.

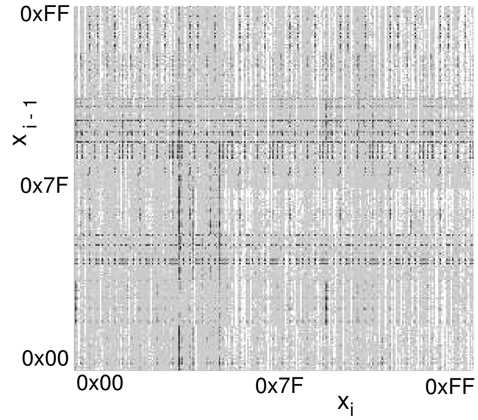
Graphical exploration of errors

We have found that individual errored octets do not appear to be clustered within frames but are independent of each other. However, we are interested in whether earlier transmitted octets have an effect on the likelihood of a subsequent octet being received in error. The encoding of an octet in 8B/10B is determined by the *running disparity* which depends on the octet sequence leading up to the current octet [3].

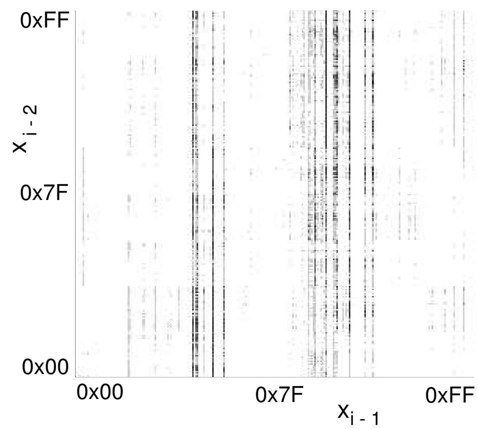
We collect statistics on how many times each transmitted octet value is received in error, and also store the sequence of octets transmitted preceding this. The error counts are stored in 2D matrices (or histograms) of size 256×256, representing each pair of octets in the sequence leading up to the errored octet: one for the errored octet and its immediate predecessor, one for the predecessor and the octet before that, and so on. We normalise the error counts for each of these histograms by dividing by the matrix representing the frequency of occurrence of this octet sequence in the original transmitted data. We then scale each histogram matrix so that the sum of all entries in each matrix is 1.

Figure 1(a) shows the error frequencies for the "current octet" X_i (the correct transmitted value of octets received in error), on the x-axis, versus the octet which was transmitted before each specific errored octet, X_{i-1}, on the y-axis. Figure 1(b) shows the preceding octet and the octet before that: X_{i-1} vs X_{i-2}. Vertical lines in Figure 1(a) are indicative of an octet that is error-prone independently of the value of the previous octet. In contrast, horizontal bands indicate a correlation of errors with the value of the previous octet; these appear as vertical lines in Figure 1(b). It can be seen from Figure 1 that while correlation between errors and the value in error, or the immediately previous value, are significant, beyond this there is no apparent correlation. The equivalent plot for X_{i-2}, X_{i-3} produces a featureless white square.

It is illustrative to consider the octets which are most subject to error, and the 8B/10B codes used to represent them. In the psuedo-random data, the following ten octets give the highest error probabilities (independent of the preceding octet value):



(a) Error counts for X_i vs. X_{i-1}

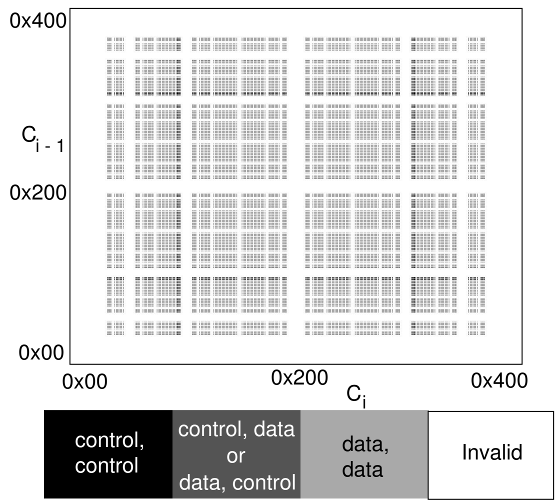


(b) Error counts for X_{i-1} vs. X_{i-2}

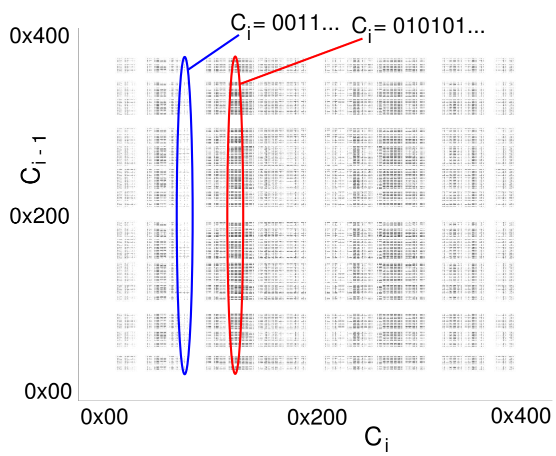
Fig. 1. Error counts for pseudo-random data octets

0x43, 0x8A, 0x4A, 0xCA, 0x6A, 0x0A, 0x6F, 0xEA, 0x59, 0x2A. It can be seen that these commonly end in A, and this causes the first 5 bits of the code-group to be `01010`. The octets not beginning with this sequence in general contain at least 4 alternating bits. Of the ten octets giving the lowest error probabilities (independent of previous octet), which are 0xAD, 0xED, 0x9D, 0xDD, 0x7D, 0x6D, 0xFD, 0x2D, 0x3D and 0x8D, the concluding D causes the code-groups to start with `0011`. Fourier Transforms (FTs) were generated for data sequences consisting of repeated instances of the code-groups of 8B/10B. Examining the FTs of the code-groups for the high error octets, the peak corresponding to the base frequency (625MHz, half the line rate) is pronounced in most cases, although there is no such feature in the FTs of the code-groups of the low error octets.

The 8B/10B codec defines both data and control encodings, and these are represented on a 1024x1024 space in Figure 2(a), which shows valid combinations of the current code-group (C_i) and the preceding one (C_{i-1}). The regions of valid and invalid code-groups are defined by the codec's use of 3B/4B



(a) Valid $C_{(i-1)}, C_i$ pairs



(b) Errors using *day-trace* as a function of code-groups

Fig. 2. The codebook for 8B/10B represented on a 1024x1024 space

and 5B/6B blocks [4].

In Figure 2(b) the octet errors found in the *day-trace* have been displayed on this codespace, showing the regions of high error concentration for real Internet data. It can be seen that these tend to be clustered and that the clusters correspond to certain features of the code-groups. Two groups of clusters have been ringed, those that are indicated as $C_i = 0011\ldots$ represent those codes with a low-error suffix. In contrast the ringed values indicated as $C_i = 010101\ldots$ indicates the error-prone symbols with a suffix of 0xA. This observed bit frequency dependent error patterning is mainly due to electrical/optical interface effects.

III. Implications

In [5] we documented the occurrence of error *hot-spots*: data and data-sequences with a higher probability of error, resulting in packets with those payloads being discarded with a higher-than-normal probability. The graphical analysis presented here is particularly useful for Gigabit Ethernet, where previous octets may affect the error probability for the current octet. However, the cause of the observed non-uniformity of errors will be common to many coding schemes.

An analysis of the contents of *day-trace* data along with other data derived as part of our network-monitoring work allows us to conclude that in addition to (user) data-payloads the error-concentrating effects will cause a significant level of loss due to the network and transport-layer header contents. In one hypothetical case, if a user was on a machine with an IP address that consisted of several high-error-rate octets their data could potentially be up to 100 times more likely to be corrupted and discarded at the Gigabit Ethernet layer.

Our investigation reveals implications for the CRC of ethernet as well as for checksums in higher layers in the protocol stack (TCP, UDP, IP). Additionally, we have started to document similar issues in Sonet-based schemes including 10 Gbps Ethernet.

IV. Conclusions.

We observe that the errors in Gigabit Ethernet in a low-power regime are not uniform as may be assumed. Examining the 8B/10B coding scheme, we have documented failures inducing, at best, poor performance and, at worst, undetected errors that may focus upon specific networks, applications and users. This content-specific effect is particularly insidious because it occurs without a total failure of the network.

Our prototype optical packet switched system illustrates how future optical networks will consist of an increasingly large number of diverse elements, with tighter limitations on the optical power budget. The design of these new networks must carefully consider the physical layer and its effects on higher level network protocols.

References

[1] D. McAuley, “Optical Local Area Network,” in *Computer Systems: Theory, Technology and Applications*, A. Herbert and K. Spärck-Jones, Eds. Springer-Verlag, Feb 2003.

[2] IEEE, “IEEE 802.3ah — Ethernet in the First Mile,” 2004, standard.

[3] ——, “IEEE 802.3z — Gigabit Ethernet,” 1998, standard.

[4] A. X. Widmer and P. A. Franaszek, “A DC-Balanced, Partitioned-Block, 8B/10B Transmission Code,” *IBM Journal of Research and Development*, vol. 27, no. 5, pp. 440–451, Sept. 1983.

[5] L. B. James *et al.*, “Structured Errors in Optical Gigabit Ethernet,” in *Passive and Active Measurement Workshop (PAM 2004)*, Apr. 2004.